Chapter

Evaluating the Processes and Procedure of Digitalization Workflow

Collence Takaingenhamo Chisita, Oluwole O. Durodolu and Joseph Ngoaketsi

Abstract

Digitisation is the practice of converting physical information into a digital (computer-readable format), by using digital technologies to modify the existing structure by enhancing the efficiency of an organisational process, foster reliability, and quality. This is a method of incorporating conventional records into a digitised form by eliminating redundancies and limiting the communications chain. This will help to improve accessibility and simplify better information exchange for users. The beginning of a digital revolution in any establishment is to appraise the manual methods with the view to improve and graduate to a user-friendly modern system. Digital workflow is a progressive, reliable arrangement of data, procedures, and responsibilities that make information is more permanent and management easy to access and enable the preservation of crucial data. This research set out to support workflow audit by revealing specific indicators to assist in processes that will enhance digital migration.

Keywords: digitization, access to information, library consortia, digital libraries, digitization workflow

1. Introduction

The essence of increasing the life span of information stored in papers containing resources of permanent value is to digitise. The pervasive and proliferating digital transformation requires organisations with the agility and the ability to react quickly to changes in the business environment [1]. Digitisation is an innovation brought about as a result of developments in computerised technology that can convert papers to an electronic format that is more permanent and devoid of environmental hazard that can eclipse the life span of information stored in the paper [2]. Theoretically, the procedure of digitisation entails converting an analogue image and other media its equivalent electronic format. In the context of this article, some of the vital issues like, the selection of documents, the scanning and image capture, arranging metadata and arranging for required hardware and software selection are essential for the development of digitisation workflow [3].

With the development of digitalization in full swing, many are pondering how the implementation of the new digital distruptive technologies will spur and inspire the creation of new jobs and destruction. The specific tasks that technologies take

on and how many new jobs the created of others. Through the adoption of new digital technologies. Some maintained that most tasks that are at risk of automation are those performed is an area for concern for policy makers in government, industry, higher education and civil society by rather low- to medium-skilled employees, while most new tasks that emerge from the adoption of digital technologies complement high-skilled labor [4].

Elaiess [5] and Arms [6] viewed a digital library as a systematically managed collection of information with allied services, where the information is stored in digital formats and reachable over the Internet. Witten and Bainbridge [7] highlighted the following as the key features of a digital library:

1. Prearranged and managed a collection of digital objects;

2. Available, accessible or obtainable over the Internet or server;

3. A universal information infrastructure; and

4. Provides service to users.

Kane et al. [8] defined digital maturity as the extent digitisation has transformed the processes, talent engagement and business models of an organisation. The indicators of including, digital maturity include;

1. The clear and coherent digital strategy incorporated, with the ability to articulate the value of digital technologies to the organisational future;

2. Comfort in taking risks and embracing failure as a prerequisite for success; and

3. Investment in organisational capabilities.

Shuva [9] argued that in the era of digital technologies, a digital library has become one of the most frequently used terms in the library and information science arena. Digital libraries have become the magic bullet to rescue governments, academic institutions, industry and users from the COVID-19 pandemic regulations which prevents physical contacts and encourages social distancing. Such techno-centric institutions refer to systems that are very heterogeneous in scope and provide different functions. These systems range from digital objects and metadata repositories, reference-linking systems, archives, and content administration systems to complex systems that integrate advanced digital systems [9]. Libraries are information systems whether traditional or modern because they collect, process, store, analyse and disseminate information for specific purposes to specific user groups [10].

Tihinen et al. [1] viewed digitalisation as one of the prominent trends transforming the information landscape, society and business in the near and long term future. Digitalisation refers to the action or process of digitising; the conversion of analogue data into digital form [1]. Digitalisation is the critical enabling issue for providing internal efficiency in organisations, or for delivering external opportunities such as new services or offerings to customers [1]. The use of various digital technologies has become a core mission of libraries and related institutions [11]. Belhi et al. [12] argued that the wide adoption of information technology obliges organisations to adapt its resources to be part of the digital era. According to the authors, the high development pace of technology has resulted in the fear of digital obsolescence as a critical factor than the fear of physical data loss. Digital

obsolescence, or data extinction, refers to the state whereby the archived data becomes no more readable or usable [12]. The files meant to be read or edited with a certain programs (e.g. Microsoft Word) might be illegible in other programs, and as operating systems and hardware changes, even old types of programs developed by the same company become difficult to use on the new platform.

Libraries as cultural institutions accommodate treasured documents, as a result, must carefully consider digitisation as a way to preserving the information resources to circumvent the loss of the originals, this is the position of modern librarians. Usually, in the library environment, digitisation comprises of scanning, photographing analogous pieces like cherished books, maps, manuscripts, correspondence, which are considered but not limited to as rare, exceptional, and tremendously delicate collections, and then transforming these resources into a digital environment where the lifespan can be permanent and the integrity preserved [13].

Digitalisation has many recognisable advantages such as instant accessibility to information, easy and speedy communication and capacity to share and exchange information, the generation of new jobs, better opportunities, and increased transparency and visibility [14]. The aim of digitisation is to improve access and advance preservation [15]. Digitalisation enables libraries and related institutions to provide virtual access to content in order to ensure the discoverability and the retrievability of the content and enhance the preservation of the content by avoiding the wear and tear of original works. The proliferation of digital technologies and the drive towards the fourth industrial revolution (4IR) should be viewed as an opportunity for libraries and related institutions to improve services to users by responding to their dynamic needs by adapting the innovative emerging digital technologies including Virtual Reality (VR).

The Digitization Workflow is an approach to explain the step by step arrangement of digitising information resources to accomplish the process of digitisation including its various phases, like the process of material selection, preparation of documents, scanning/OCR use scanner, processing for editing, quality assurance, metadata and indexing, back-up and archiving, publishing in a digital repository and finally checking out. The rationale of this workflow is to provide history tracking of actions and flexibility to accomplish multiple projects with multiplicity and diversity of materials concurrently. Additionally, it supports the convinient assimilation of tools used to implement the functions of the workflow. Optical Character Recognition (OCR) is a technology that allows the conversion of diverse types of the documents and materials, such as PDF files or images captured by the use of a digital camera into editable and searchable data formats [16].

The first stage of digitization is document selection which requires expertise in knowing which documents to include and for what reasons. The preparation of document is key to enable selected documents a flawless procedure that is devoid of office objects such as: document clips, sticky notes, pins and spiral bind. Scanning/OCR (Optical Character Recognition) is the conversion of physical documents to electronic format which requires the use of highly efficient, reliable and speedy electronic scanning machine. Optical Character Recognition is used in converting images, handwritten or printed text into machine-encoded text, whether from a scanned document, this is important because it will make the document searchable and easy to locate and retrive from online. The integrity of the document it is of paramount importance in order to ensure quality assurance of the document by properly editing the information content so that the document will be free from both content and grammatical errors. The final stage of the digitization process is to ensure back-up and archiving this will guarantee the security of the document and information content, in case of hacking, social engineering, cross-site scripts virus attack or other unforeseen problem that might compromise the integrity of the document [17]. These are the step by step

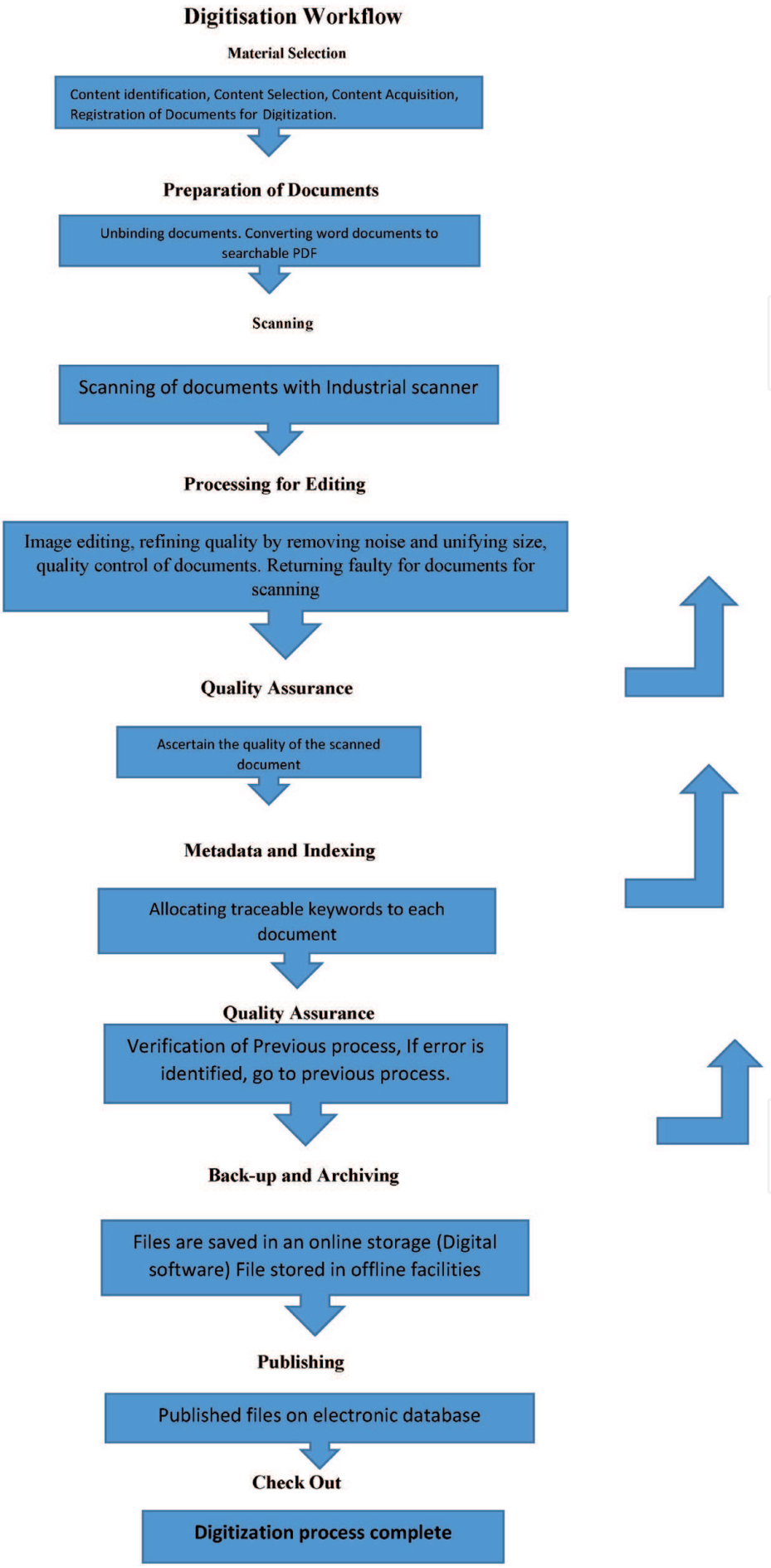

Figure 1.
Digitalisation workflow chart.

procedure of digitization workflow before finally publishing the content to make it visible within the local network and its availability and accessibility on the internet for the benefit of the general public (**Figure 1**).

The Digitization Workflow is an approach to explain the step by step arrangement of digitising information resources to accomplish the process of digitisation including its various phases, like the process of material selection, preparation of documents, scanning/OCR use scanner, processing for editing, quality assurance, metadata and indexing, back-up and archiving, publishing in a digital repository and finally checking out. The rationale of this workflow is to provide history tracking of actions and flexibility to accomplish multiple projects with multiplicity and diversity of materials concurrently. Additionally, it supports easy assimilation of tools used to implement functions of the workflow. Optical Character Recognition (OCR) is a technology that allows the conversion of diverse types of documents and materials, such as PDF files or images captured by the use of a digital camera into editable and searchable data format [16].

Abollado et al. [18] noted that while digital workflows are a suitable solution for managing business processes complexity in the engineering industry, it is equally essential to be aware of all the challenges that are associated with implementing a workflow system. Such grandiose projects require concrete management engagement, end-user involvement, tools and system integration and a sensible implementation plan. Abollado et al. [18] observations were drawn from the engineering sector, but they apply to library science with regards to the ongoing digitisation projects.

Digital workflow projects underpin organisational performance. The previous phases of a workflow project are generally more critical to the overall success of the project. There are several recommended practices for implementing a successful workflow management system in an aerospace company, for example, obtaining the support of senior management, integrating the digital workflow with current techniques and securing the support of end-users who are also key stakeholders [18].

The digital workflows and workflow management tools offer an opportunity to improve, automate and streamline the underlying processes in any business, including the library [19]. Such projects help to enhance communications, among others, benefit—performance, accountability and visibility. The essential criteria for selection for digitisation are the copyright status of the original materials. Images are supposed to have their copyright held by the establishment. If the organisation does not have the right to digitise, then other images must be chosen, or the project cannot proceed for fear of litigation. Preparation of the digitisation process includes reviewing the procedures in place from the beginning to the end of the process of production facility up to the final delivery digital images and data back. Other important aspects of the digitisation workflow includes generating indexing and keyword to includes convinient information accessibility and precision. It is vital to ensure quality control so that the output will be acceptable to the general public and meet international standards. To securing data and enable the integrity of the information is vital, data encrypted and securely kept to guard against loss of information [20].

2. Digital preservation challenges

Content digitalization has becoming an established enterprise in the information age, giving birth too many innovative business ideas, models and much more. This has led to numerous disciplines and industries taking advantage of its opportunities and broadening and widening access to organisational resources. As a result of several advances in information and communication technologies many organizations have adapted to digital models driving the new digital era.

The general misunderstanding is that to digitize a material is the same as digital preservation. To digitize material is the act of converting something from an analog to a digital format. For instance, scanning a photograph and retaining a digital copy on a computer. This is basically the first step in digital preservation. To digitally preserve a material is to prolong the content over a long time [21].

Several galleries, libraries, archives, and museums (GLAMs), and other cultural institutions, undergo uphill task catching up and up to date in digital preservation. Digitization is a onerous and time-consuming activities, mostly because it is contingent on the critical condition of the holdings previous to the time of digitized. The possibility exist that the materials may be fragile and delicate that if care is not taken the information resources may be lost or become damaged irreparably; light emanating from the scanner can destroy old photographs papers and documents. Notwithstanding the potential damage, one rationale for content digitization is as a result of constant use of a material, therefore, digitization will preserve the original copy from total degeneration [22].

The process of digitization is relatively expensive. Organizations ensure best quality in digital copies by retaining the best possible. Raising money to meet the expense required for the equipment is another major challenge. The quality of personnel may also limit the process of digitization. Archivists and librarians must be familiar with the aspirations and desires of their patrons and try to prioritize and meet the needs digitally [22].

3. IFLA guideline for digitization

The International Federation of Library Associations and Institutions (IFLA, 2000) [23] a foremost international body representing the interests of libraries and information professionals, in a report in 2000 identified guidelines for digital project, which has become popular in setting standard for the implementation and practice of digitization. Regardless of this standard, studies of digitization shows that it is rarely utilized as standards for evaluation [24].

4. Process of selection for digitization

The formulation of selection policies is a fundamental element of digital schemes, and various selection guidelines and standards have been established by organizations to authenticate their selection processes for digitization regarding external benchmarks, particularly with the growth of partnerships for digital projects. It is important to prioritize the materials to digitize based on the need of the patron, usage of information material is an alternative factor which helps to decide a collection's priority for digitization [25].

5. Technical requirements and implementation

The technical Standards for digitization requires to sourcing records in line with best practice and technical standards when the digitized record is projected to stand in place of the source record as the official record of information resources [26].

There are steps to follow in the implementation of digitization projects, determining the vision for implementing digital transformation, analyzing the market by determining the needs of the patron because of quick technology improvement and industry modifications, are crucial for creating the appropriate and up-to-date

strategy. Another step involves assessing the a current standing which is a way to evaluate what technology to update and to identify what digital tools to change will help in prioritizing the efforts and investment in digital transformation strategy in the best possible way [27].

6. Using the system development life cycle for understanding of digitization process

Elaiess [28] and Elaiess [5] viewed the standard development life-cycle (SDLC) as a means to provide the standard methodology and high-level operational guidelines within which software is developed and maintained. It is the process of understanding how an information system can support business needs, designing the system, building it, and delivering it to users. Cervone [29] argued that the system development life-cycle concept had been applied mainly to system development projects for years. Project teams developing digital library systems can be more effective if they understand the expectations and outcomes of each phase of the system development life-cycle [29]. Systems development is the art and science of creating human made information systems to satisfy predetermined needs. The systems development life-cycle can be used for the traditional and the modern library materials with regards to creation, processing, dissemination and preservation. It is a problem-solving process integrates appropriate elements of humankind's knowledge base to create new knowledge specific to the problem and, as a result, define a solution to the problem. Grenci and Hull [30] highlighted that Information systems development methodologies (ISDM) encompass the sum total of methods that are used for developing and implementing information systems applications. The use of information systems in the organisational environment has been growing in recent times, and this justifies the need for broadening and widening understanding of how such systems operate and how they can benefit libraries and related institutions.

Duarte and Costa [31] viewed Information Systems as important tools to enhance the efficient management of information and other knowledge assets in organisations. The life cycle model is very important because each phase of the process influences the phase after, which means that phases of the life cycle have a great influence on the global success of these systems. Duarte and Costa [31] argued that it was vital to know all the process and its critical success factors to make sure that there are no mistakes. The systems development life cycle must be understood because it underpins the success of a library digitization project. The digitization of libraries is now on the agenda of many academic institutions and governments especially now in the era of the COVID-19 pandemic. The aforementioned processes and procedures of digitalization workflow are useful for the development of digital libraries in the era of the COVID-19 pandemic.

7. Conclusion

This article presents a substantial academic work highlighting different phases of digitization workflow and the process of electronic or digital conversion of documents. Digitization is important because it helps to enhance data processing, improving data storage, fast-tracking transmission and improving efficient service delivery. It also facilitates data sharing and retrieval, and it has proven to be the most satisfactory way of preserving information for a considerable length of time. This research also simplifies the chain of processes and tasks involved in a

digitization project, thus facilitating uniformity and dependable results relating to the digitisation of a large volume of objects. It should be noted that the documentation of the workflow is critical for tracking material within the context of a series of the stages of the process and identifying technical glitches. This article also highlighted some of the challenges of digitization. For instance, damages caused by environmental problems like, earthquake, rainfall, humidity, fire, and other human factors like careless handling, defacement and inappropriate support during storage and cyber threats. The deterioration of library resources like books, journals, and other materials forms the fundamental challenges of libraries which makes preservation and conservation imperative. Finally, this study also revealed the digital lifecycle and its phases through which digitization process go through to attain full conversion and this is the passage of modernization that pursues constant regulations to attain innovative expectations.

Author details

Collence Takaingenhamo Chisita*, Oluwole O. Durodolu and Joseph Ngoaketsi
Department of Information Science, University of South Africa, Pretoria, South Africa

*Address all correspondence to: chisitacollence@gmail.com

References

[1] Tihinen, M., Kääriäinen, J., Ailisto, H., Komi, M., Parviainen, P., Tanner, H., Tuikka, T. and Valtanen, K., (2016). The Industrial Internet in Finland: on route to success? (Eds.), The Industrial Internet in Finland: on route to success?. Espoo, Finland: VTT, VTT Technology; 278. ISBN 978-951-38-8484-0, http://www.vtt.fi/inf/pdf/technology/2016/T278 .pdf, 2016. Accessed 8 November, 2020.

[2] Morrison, et al. (n d), Guide to Creating and Documenting Electronic Texts. Retrieved on 16/10/2020, ota.adhs.ac.uk/documents/creating/Chap.1.html

[3] Swain, D.K and Panda, K.C. (2009). An Overview of Digitization of Information Resources. Retrieved on 16/10/2020, file:///C:/Users/USER/Downloads/An_Overview_of_Digitization_of_Information_Resourc.pdf

[4] Balsmeier, B and Woerter, M (2019). Is this time different? How digitalization influences job creation and destruction. Volume 48, Issue 8.

[5] Elaiess, R. (2017). Design and implementation option for digital library systems. International Journal of Systems Science and Applied Mathematics, *2*(3), 70-74.

[6] Arms, (2000). Digital libraries. Cambridge: MIT Press.

[7] Witten, I. and Bainbridge, D., (2003). How to build a digital library. San Francisco: Morgan Kaufmann.

[8] Kane, G. C., Palmer, D., Phillips, A. N., Kiron, D., & Buckley, N. (2015). Strategy, not technology, drives digital transformation. MIT Sloan Management Review and Deloitte University Press, *14*(1-25).

[9] Shuva, N. Z. (2012). Building digital libraries in Bangladesh: A developing country perspective. The International Information & Library Review, *44*(3), 132-146.

[10] Rainer, R. K., Prince, B., Splettstoesser-Hogeterp, I., Sanchez-Rodriguez, C., & Ebrahimi, S. *Introduction to information systems*. John Wiley & Sons, 2020.

[11] Hughes, L. M. (2004). *Digitising collections: strategic issues for the information manager* (Vol. 2). Facet Publishing.

[12] Belhi, A., Foufou, S., Bouras, A., & Sadka, A. H. (2017). Digitisation and preservation of cultural heritage products. In *IFIP International Conference on Product Lifecycle Management* (pp. 241-253). Springer, Cham.

[13] Deac, V., Burz, I., & Deac, A. (2019). Evaluation and Protection of Library Collections with Heritage Value. Establishing Ante-Digitization and Post-Digitization Measures. Romanian Journal of Library & Information Science / Revista Româna de Biblioteconomie Si Stiinta Informarii, *15*(2), 51-60. https://doi.org/10.26660/rrbsi.2019.15.2.51

[14] Durodolu, O. O. (2007). 'The Relevance of Information Technologies to Libraries as Information Centres: A Survey of two special Libraries' International Journal of Information and Communication Technology (ICT) Vol. 4. Number 2, Pp. 90 -104

[15] Bandi, Angadi and Shivarama (2015). Best practices in Digitisation: Planning and workflow processes. Emerging Technoology and Future of Libraries: Issues and challenges.

[16] Ray Smith (2007). An Overview of the Tesseract OCR Engine. Retrieved

May 23, 2020. https://web.archive.org/web/20100928052954/http://tesseractocr.googlecode.com/svn/trunk/doc/tesseracticdar2007.pdf

[17] Adeleke, A, Durodolu, O. O. [and] Adekanye, E. A. (2011). 'Internet use and academic integrity among Education undergraduate in Nigeria' Library Progress (International) Vol 31 (2) (International).

[18] Abollado, J. R., Shehab, E., & Bamforth, P. (2017). Challenges and Benefits of Digital Workflow Implementation in Aerospace Manufacturing Engineering. Procedia CIRP, *60*, 80-85.

[19] National Archives, National Historical and Records Commission. digitisation Workflow https://www.wisconsinhistory.org/pdfs/la/Digitization-State/7_Digitization-Workflow.pdf Accessed 16 April, 2021

[20] Taube, M (2017). 10 Steps to Document Scanning Happiness. Retrieved on 16/10/2020, https://www.digiscribe.info/blog/how-to-measure-document-scanning-quality/

[21] Chou, R (2011), why digitization is not digital preservation. Retrieved on 24/12/2020 https://anthroarchivist.wordpress.com/2011/10/07/why-digitization-is-not-digital-preservation/

[22] American Library Association (2008). What Is Digital Preservation? Retrieved on 24/12/2020 https://journals.ala.org/index.php/ltr/article/view/4224/4808

[23] International Federation of Library Associations (IFLA) (2004). Guidelines for digitization projects for collections and holdings in the public domain, particularly those held by libraries and archives. Retrieved December 28, 2020, from https://www.ifla.org/files/assets/preservation-and-conservation/publications/digitization-projects-guidelines.pdf

[24] International Federation of Library Associations (IFLA) (2005). Digital reference guidelines. Retrieved January 5, 2006, from http://www.ifla.org/VII/s36/pubs/drg03.htm

[25] Lucidea (2018) Selection for Digitization – Best Practices. Retrieved on 16/10/2020, https://lucidea.com/blog/selection-for-digitization/

[26] Government of Western Australia (2014). Technical Standards for Digitization and Disposal of Patient Records Policy Toolkit December 2014. Department of Health

[27] Zhovtyuk, V (2018). How to Develop a Digital Transformation Strategy: Step-by-Step Guide. Retrieved on 16/10/2020, https://medium.com/@VitaliyZhovtyuk/how-to-develop-a-digital-transformation-strategy-step-by-step-guide-f8e4bd5719b6

[28] Elaiess R., (2009) General guidelines for designing bilingual low-cost digital library services suitable for special library users in developing countries and the Arabic speaking world. In: Proceedings of the world library and information congress: 75th IFLA general conference. Milan, Italy 23-27 August 2009. [Online] Accessed 24 November, 2020.

[29] Cervone, H.F. (2007), "The system development life cycle and digital library development", OCLC Systems & Services: International digital library perspectives, Vol. 23 No. 4, pp. 348-352. https://doi.org/10.1108/10650750710831484

[30] Grenci, R. T., & Hull, B. Z. (2020). New dog, old tricks: ERP and the

systems development life cycle. Journal of Information Systems Education, 15(3), 7.

[31] Duarte, A. I. M., & Costa, C. J. (2012, June). Information systems: Life cycle and success. In *Proceedings of the Workshop on information systems and design of* communication (pp. 25-30).